115th MPEG Geneva (CH), Switzerland, 30 May - 3 June 2016, Meeting Report

Panos Kudumakis, Krishna Chandramouli and Ebroul Izquierdo
qMedia, Queen Mary University of London

Contents

1 Internet of Media-Things and Wearables (IoMTW)

Today, "Internet of Things" (IoT) is an idiom currently encompassing a large variety of research, development and market efforts related to the communication between smart objects. While such an definition is, by its very nature, quite fuzzy, the market reality is very clear: the number of devices connected to the Internet will reach to 50 billion in 2020 and, together with the data they are producing, they will result in the first source of wealth in the world, at least for the IT industry.

An important factor contributing to the growing adoption of IoT (Internet of Things) is the emergence of wearable devices, a category with high growth market potential. Wearable devices, as understood commonly, are devices that can be worn by or embedded in a person and have the capability to connect and communicate to the network either directly through embedded wireless connectivity or through another device (e.g. smartphone) using Wi-Fi, Bluetooth, or another technology.

After a preliminary study, MPEG decided to align the standardization activities related to IoT and wearable devices inside a unique effort, referred to as *Internet of Media-Things and Wearables*.

At the 115th meeting, MPEG produced *draft* documents of Requirements (N16346) and Use Cases (N16345) as well as a *draft* Call for Proposals on Internet of Media-Things and Wearables (IoMTW). A final Call for Proposals will be issued at the 116th MPEG meeting.

Preliminary use cases addressing the following areas:

Smart spaces: Monitoring and control with network of audio-video cameras

- Face recognition to evoke sensorial actuations
- Human tracking with multiple network cameras
- Networked digital signs for customized advertisement
- Intelligent firefighting with IP surveillance cameras
- Automatic video clip generation by detecting event information

Smart spaces: Navigation

- Blind person assistant system
- Personalized navigation by visual communication
- Personalized tourist navigation with natural language functionalities

Smart environments

- Smart factory: Car maintenance assistance A/V system using smart glasses
- Smart museum: Augmented visit museum using smart glasses
- Smart car: Head-light adjustment and speed monitoring
- Smart house: Light control and vibrating subtitle

Smart collaborative health

- Increasing patient autonomy by remote control of left-ventricular assist devices
- Diabetic coma prevention by monitoring networks of in-body / near body sensors
- Enhanced physical activity with smart fabrics networks

General requirements
1. Shall be able to communicate with other things
2. Shall be able to be configured dynamically
3. Shall be able to provide a unique identifier
4. Shall be able to provide characteristic descriptions (e.g., capabilities, current status)
5. Shall be able to send context information to other things
6. Shall be able to be discovered
7. Shall be able to support security
8. Shall be able to expose information related to its reliability
9. Shall be able to support timeliness aspects
10. Shall be able to have (limited) storage capabilities

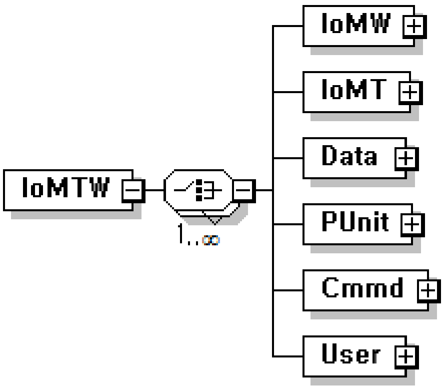

Figure 1- Internet of Media-Things and Wearables (IoMTW) root elements.

Output Documents
N16345 - Use cases for Internet of Media-Things and Wearables
N16346 - Requirements on Internet of Media-Things and Wearables
N16347 - Draft Call for Proposals on Internet of Media-Things and Wearables
N16348 - Status of discussion on IoMTW

N16349 - Liaison Statement to JTC 1/WG 10 on IoT

2 Visual Identity Management AF (VIM-AF)

Citizens are voluntarily sharing significant amounts of information (data, pictures and video) about themselves and their activities through social networking applications, which are being commercially exploited and traded for marketing purposes, but also used for malicious activity.

In those social networks, access control is provided and done by the social networks themselves without any control by the user. To restore citizens' confidence in online data collection practices, submitted media should be encrypted by the user and can only be viewed according to user wishes such as: by a group of friends, purpose of sharing, time, date or metadata.

In order to provide user control over processing and sharing of multimedia content, a flexible, effective and scalable mechanism is to provide users a way to express their control desires in a form that can be processed and monitored systematically, consistently and persistently throughout the lifecycle of the multimedia content. These control desires can be considered as multimedia usage-control policies, much like data usage-control policies many websites (e.g., Facebook) use to protect user identity and data usage-control. *The difference here is that, in the user control case, it is the users, not the websites, who specifies how their multimedia content should be processed and shared.*

In order to express users' desire on controlling usage of their multimedia content, the following should be supported in a usage-control policy expression language such as a profile of the MPEG-21 CEL:

- Specification of ownership of the multimedia content and their components, with optional authentication checking mechanisms.
- What kinds of processing, or types of operations, can be applied to the multimedia content, and their deontic characteristics such as permissions, obligations and prohibitions (i.e., rights, duties and bans in the MPEG-21 CEL terms).
- How to share, with whom, for how long and under what conditions, also along with their deontic characteristics such as permissions, obligations and prohibitions (i.e., rights, duties and bans in the MPEG-21 CEL terms).

System for image identity privacy

Figure 2 illustrates a conceptual framework for managing privacy of users when pictures are taken and being shared among users.

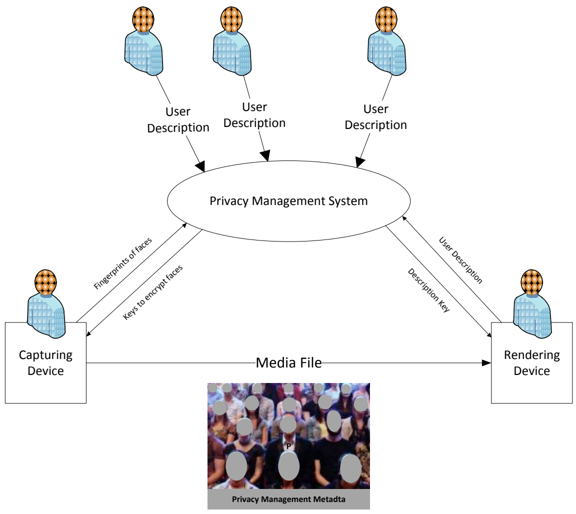

Figure 2 - Framework for privacy management of pictures.

- Each user registers his privacy management metadata to a Privacy Management System (PMS), containing the following information
    - Unique ID of the user
    - A fingerprint of the face of the user
    - A list of the IDs of the users who are allowed to see his face
- The PMS automatically generates encryption keys for each registered user's face and shares the decryption keys with the users who are allowed to see his face
- When a picture is taken by a capturing device, a unique ID of the picture is generated and it is provided to PMS with the list of fingerprints of the faces in that picture with sequential numbers associated for every face in the picture
- The PMS identifies the list of users in the picture by matching the fingerprints received and the fingerprints of the registered users and provides appropriate encryption keys for each face to the capturing device
- The capturing device creates a file of the picture according to the user's privacy management metadata capturing the picture, including the following information
    - Unique ID of the picture
    - List of sequential numbers of encrypted faces and their location and size
- The capturing device shares the file of the picture to a rendering device
- A rendering device tries to decrypt the faces by using the privacy management metadata of the user trying to see the picture and the keys received from the PMS; in doing so
- The rendering device sends the following information to the PMS
    - The ID of the picture the user is trying to see
    - The description of the user trying to see the picture
    - The context of the user
- The PMS evaluates whether the user trying to see the picture is listed as one of the users who can see the faces of the users in the picture. If the user trying to see the picture is not listed as the one who can see some of the faces of the users in the picture, then the PMS sends to those

users a description about the user trying to see the picture. If those users agree to show their faces to the user trying to see the picture then the PMS sends to the rendering device the keys to decrypt their faces.

Requirements

Visual Identity Management Application Format SHALL:

- Support user's full control of privacy on media where his/her trace is recorded
- Support user's full control of privacy on any type of media (e.g., video and still pictures)
- Support partial (e.g., location, size) en/decryption on media to secure user privacy without causing crashes in the video decoder
- Support selective (e.g., user specific) en/decryption on media to secure user privacy
- Support exchange of privacy description information between PMS, User, Capture device and Rendering devices, such as:
    - Unique ID of the user
    - A fingerprint of the face of the user
    - A list of the IDs of the users who are allowed to see his face
    - The ID of the picture the user is trying to see
    - The description of the user trying to see the picture
    - The context of the user

Considered technologies for Visual Identity Management AF

Fingerprint for a face of user

ISO/IEC 15938-13:2015 Compact descriptors for visual search (CDVS) shall be used to represent fingerprint of the face of a user in a picture. For video, CDVA might be used.

Privacy description of user

ISO/IEC 21000-22 User Description with some proposed extensions shall be used to represent privacy description of a user.

Region encryption

ISO/IEC 23001-7 Common Encryption with some proposed extensions shall be used to indicate to the video decoder that the encrypted bit stream can be decoded even if it is considered as "encrypted".

Support of AVC and HEVC backward compliant decoder

ISO/IEC 14496-10 (AVC) and ISO/IEC 23008-2 (HEVC) legacy decoders shall be able to decode bit streams that carry Visual Identity Management AF functionality, i.e. non-protected bit streams parts should be processable without causing legacy codecs to crash.

Output Documents

N16226 - Use cases and requirements on Visual Identity Management AF

3 A new MPEG-21 UD use case: Loudness Control on Smart Devices

According to psychoacoustic experiments, it is known that human imagines high sound as high quality. Such a human's characteristic is exploited by sound producers and providers by competitively raising sound pressure levels to make profits as shown in Figure 3.

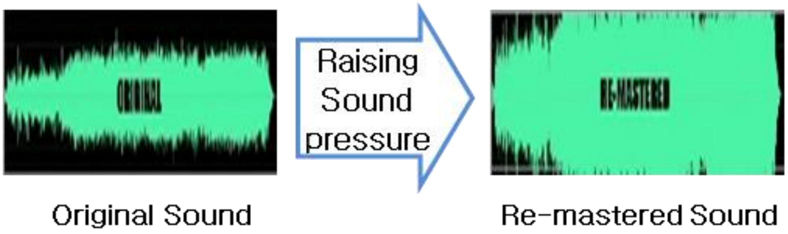

Figure 3 - Raising sound pressure.

Echo Nest (American Music Research Institute) analyzed popular top 5,000 songs since 1950. As a result, the "Loudness Average" has continuously increased every year, especially, it has increased 39% during recent two decades. Accordingly, from 2010 to 2014, deafness increased by 5.3% and related medical expenses by 12.5%.

Thus, loudness control in smart devices is very important to protect to user's capability of hearing. Nevertheless, it is not easy to compel lowering sound pressure, while damaging provider's profit. MPEG-21 UD can help this situation by properly gathering and recommending information. User Description (UD) gathers information on user's hearing capability and preference. Service Description (SD) provides level of sound and Context Description (CD) gathers information on level of peripheral noise and device capability. Recommendation Description (RD) estimates proper level of sound. The use case for Loudness Control is given in the following steps.

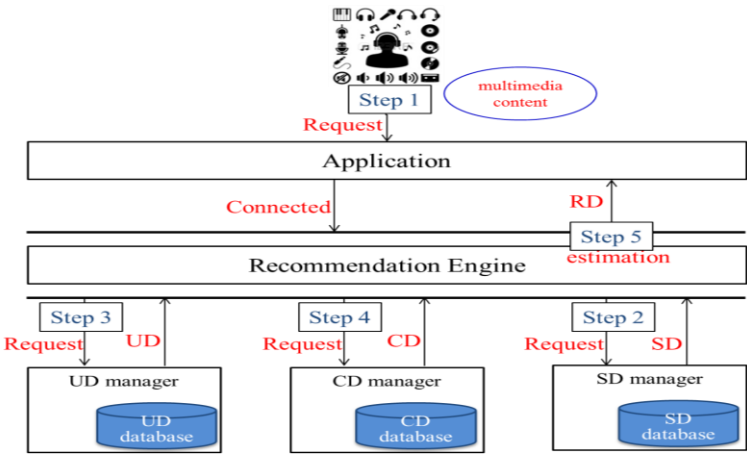

Figure 4 - Loudness control use case scenario onto MPEG-21 UD architecture.

Step 1. A user requests multimedia content to an application which is connected to a Recommendation Engine.

Step 2. The Engine seeks proper Service Providers and requests their SDs.

Step 3. The Engine requests the User's capability of hearing and preference of the content at this moment to the UD manager.

Step 4. The Engine requests level of peripheral noise and capability of the user's device to the CD manager.

Step 5. The Engine estimates a proper level of sound for the user at a given moment and recommend it to the application.

Step 6. Whenever the user moves a place to another place - a different CD comes into the Engine, a different level of sound is recommended.

Output Documents

N16315 - Technologies under consideration for extension of MPEG-21 UD

N16180 - Text of ISO/IEC 21000-22 DAM 1 Reference Software for MPEG-21 User Description

4 Media Orchestration

With the abundance of capture and display devices, and with applications and services moving towards a more immersive experience, we need the tools to be able to manage multiple, heterogeneous devices over multiple, heterogeneous networks, to create a single experience. We call this process Media Orchestration: orchestrating devices, media streams and resources to create such an experience.

Media orchestration:

- Applies to capture as well as consumption;
- Applies to fully offline use cases as well as network-supported use, with dynamic availability of network resources;
- Applies to real-time use as well as media created for later consumption;
- Applies to entertainment, but also communication, infotainment, education and professional services;
- Concerns temporal (synchronization) as well as spatial orchestration;
- Concerns situations with multiple sensors ("Sources") as well as multiple rendering devices ("Sinks"), including one-to-many and many-to-one scenarios;
- Concerns situations with a single user as well as with multiple (simultaneous) users, and potentially even cases were the "user" is a machine, although this is not yet represented in the use cases. This may have a relation with the notion of "Media Internet of Things" that is also discussed in MPEG.

MPEG issued a Call for Proposals (CfP) on Media Orchestration at its 114th meeting. The CfP resulted in 4 responses at the 115th meeting, which were reviewed and evaluated. These are briefly reviewed below:

- qMedia::MMV proposed a solution in response to CfP for Media Orchestration (m38557) related to security domain which resulted in discussion on place holders in the WD for AV patterns and time stamped metadata.
- Another proposal suggested stitching images to create a UWV media experience represented by using DASH Spatial Relationship Descriptors and MMT Composition Information (m38788).
- Another submission (m38564) proposed to use MPEG-V orientation/position as timed metadata for signalling sensor capabilities; use of ISOBMFF and MPEG-2 TS to carry timed metadata; and use of MPEG-2 TEMI for the association and correlation of multiple streams of timed data.
- Last but not least the use DVB-CSS proposed for wall-clocks synchronisation and presentation timeline of Sources and Sinks as well as the use of DVB-CSS in combination with TEMI for association and correlation of media streams (m38227).

It was decided that a second CfP would be useful with responses due for the 116th meeting. So an updated second CfP issued based on the original CfP.

Output Documents
N16336 - Evaluation Report for CfP on Media Orchstration
N16337 - Media Orchestration Requirements v.4
N16338 - Second Call for Proposals on Media Orchestration Technologies
N16194 - Working Draft of ISO/IEC 23001-13 Media Orchestration (MORE)
N16195 - Technologies under Consideration for ISO/IEC 23001-13 Media Orchestration
N16314 - Considerations on reference software for media orchestration

5 MPEG-21 Media Value Chain Ontology (MVCO) - Audio Extensions

Widespread adoption of multi-track formats such as the MPEG-A: Interactive Music Application Format (ISO/IEC 23000-12:2010) raises the issue of rights monitoring for fair and transparent royalties payment with respect to reusable tracks or even segments of them in derivative new works. This MVCO amendment for IP Entities in the audio domain addresses this issue by facilitating complex matrix based rights monitoring on time vs tracks throughout the media value chain. It defines concepts for the representation of time segments and tracks of multi-track audio IP entities. Segments and tracks may contain IP entities that can be treated as conventional IP entities as defined in MVCO. The introduction of an additional Action 'ReuseIPEntity' in MVCO enables granting permissions for the reuse of existing IP Entities in order to create new derivative composite IP Entities.

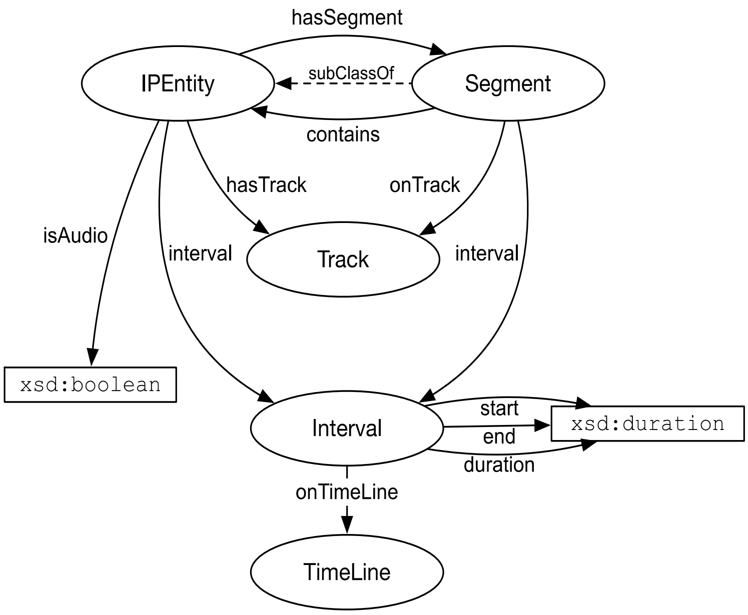

Figure 19 - Classes and relationships for the representation of IPEntity that contain other existing IP Entities. Segments may also be associated with individual Tracks of a Multi-track IP Entity.

With qMedia::C4DM proposed MVCO Audio Extensions, it is now possible to query for information about user collectives and the components of the composite IP Entities.

List members of a user collective:

```
$ java -jar rvdac.jar -r -lcu Performers
RVD Administration Console
Guitarist
Vocalist
```

List components of a composite IP Entity (including locations specified by segment and tracks where applicable):

```
$ java -jar rvdac.jar -r -lic MusicInstance
RVD Administration Console
LyricsInstance | segment: 30s to 150s | track: 2
GuitarInstance | track: 1
```

Further information on this particular use case can be found in the MVCO Audio Extensions guidelines document.

Output documents

N16177 - Request of ISO/IEC 21000-19:2010 AMD 1 Extensions on Time Segments and Multi-Track Audio

N16178 - Text of ISO/IEC 21000-19:2010 PDAM 1 Extensions on Time Segments and Multi-Track Audio

6 Point Cloud Compression

The aim of this future standard is the coding of content represented by static and/or time-varying point clouds. Recently, the investigation of new coding tools for static and dynamic 3D point clouds have shown to improve the coding efficiency with respect to existing standard solutions and new tools have been proposed and discussed within MPEG.

At the 115th meeting, MPEG produced *draft* documents of Requirements, Use Cases, Evaluation Criteria and a Dataset Description as well as a *draft* Call for Proposals on Point Cloud Compression (PCC). A final Call for Proposals will be issued at the 116th MPEG meeting.

Point Cloud Compression (PCC) use cases currently include: 3D immersive telepresence, 3D sports broadcasting, geographic information systems and cultural heritage.

General Requirements

3D Point Cloud Representation

- The 3D point cloud representation shall include 3D position (X, Y, Z) with a specification of its precision and dynamic range.
- The 3D point cloud representation shall support multiple attributes being associated with each 3D position including colour, reflectance, normal vectors and transparency or other attributes.
- The 3D point cloud representation shall support generic attributes being associated with each 3D position.
- The 3D point cloud representation shall support view-dependent attributes being associated with each 3D position.
- The representation shall support time-varying point clouds.

3D Point Cloud Compression

- Lossy compression: parameter control of the bitrate shall be supported.
- Lossless compression: the reconstructed data shall be mathematically identical to the original one.
- Temporal variations of point clouds shall be supported.
- Progressive and/or scalable coding: it shall be possible to first decode a coarse point cloud and then refine it.
- View-dependent decoding, spatial random access: it shall be possible to first decode the point-cloud corresponding to a region.
- Temporal random access shall be possible.
- Error resilience: it should be possible to cope with packet loss without having to retransmit the entire point cloud.
- Compression shall support encoding and decoding with low complexity, low latency and/or real-time implementation
- Compression should enable parallel encoding and decoding.
- Compression shall target compression rates of 10 times for lossy coding and upto 40 times for lossy coding.
- Output bitrates of 5 Mbps, 10 Mbps, 20 Mbps and 40 Mbps shall be supported

qMedia::MMV submitted a dataset consisted of human subjects captured in 3D environment where a single Kinect version 1 is used to scan the subject standing on a turntable in the center in five different poses. Each human subject is represented using ~100K points and for each human subject 5 frames of data is made available.

Output Documents
N16330 - Requirements for Point Cloud Compression
N16331 - Use cases for Point Cloud Compression
N16332 - Evaluation criteria for Point Cloud Compression
N16333 - Draft dataset description for Point Cloud Compression
N16334 - Draft CfP for Point Cloud Compression